



March 31, 2025

VIA ECF

Hon. Ona T. Wang
Daniel Patrick Moynihan
United States Courthouse
500 Pearl St.
New York, NY 10007

RE: *Authors Guild v. OpenAI Inc.*, (No. 1:23-cv-08292-SHS)
Alter v. OpenAI Inc., (No. 1:23-cv-10211-SHS)

Dear Judge Wang:

“The purpose of Rule 36(a) is to narrow the issues for trial to those which are genuinely contested.” *Azkour v. Haouzi*, 2014 WL 4467897, at *10 (S.D.N.Y. Sept. 9, 2014) (citation omitted). OpenAI refuses to admit or deny whether specific works were included in OpenAI’s training dataset or used to train OpenAI’s models. Rather than answer these questions, OpenAI has impermissibly misconstrued Plaintiffs’ RFAs and vaguely answered that [REDACTED]

[REDACTED] Ex. 1; *see also Herrera v. Scully*, 143 F.R.D. 545, 549 (S.D.N.Y. 1992) (deeming evasive denials to RFAs as admissions). OpenAI argues that it cannot respond to Plaintiffs’ RFAs using the detailed information that Plaintiffs have given OpenAI because, *inter alia*, [REDACTED]

Pursuant to Rule II(b) of Your Honor’s Individual Practices, Plaintiffs seek a conference on OpenAI’s supplemental RFA responses.

Background. As the Court is aware, well over a year ago, Plaintiffs served RFAs asking OpenAI to admit that Plaintiffs’ works were (1) included in the dataset used to train OpenAI’s models; (2) used to train OpenAI’s models; or (3) not included in the dataset used to train OpenAI’s models. OpenAI refused to respond to these RFAs, and Plaintiffs moved to compel OpenAI’s responses on February 23, 2024. *Authors Guild*, Dkt. 78. The Court heard oral argument at the September 12, 2024 conference. The Court deferred ruling on Plaintiffs’ motion and instructed the parties to confer on what Plaintiffs want OpenAI to search for and how Plaintiffs “want the search to progress.” 9/12/24 Hearing Tr. 14:14-16.

Plaintiffs inspected OpenAI’s training data in November. [REDACTED]

[REDACTED]. In December, Plaintiffs sent OpenAI a spreadsheet containing (1) the subset of works that Plaintiffs wanted OpenAI to search for and [REDACTED] At OpenAI’s

March 31, 2025

Page 2

request, Plaintiffs also gave OpenAI the Bates numbers for each of the relevant works that Plaintiffs produced.

After more than a year of refusing to respond to any of Plaintiffs' RFAs, on February 26, 2025, OpenAI agreed to respond to 290 RFAs corresponding to the subset of works in Plaintiffs' spreadsheet. OpenAI served its supplemental responses on March 27, 2025, lodging a host of brand-new objections to year-old RFAs and stating for each RFA [REDACTED]

[REDACTED] *See, e.g.*, Ex. 1 at 5. The parties conferred on these responses on March 28 and are at impasse.

OpenAI has waived its new objections. OpenAI's March 27, 2025 Supplemental RFA Responses raise new objections to Plaintiffs' December 2023 and January 2024 RFAs. For example, OpenAI now objects to [REDACTED]

[REDACTED] As an initial matter, because OpenAI could have raised each of these objections in its initial responses and failed to do so, OpenAI has waived these objections. *Banco Cent. De Paraguay v. Paraguay Humanitarian Found., Inc.*, 2006 WL 3456521, at *9 (S.D.N.Y. Nov. 30, 2006) ("The law is well settled that a failure to assert objections to a discovery request in a timely manner operates as a waiver."). The same is true of OpenAI's objections that Plaintiffs have not [REDACTED]

[REDACTED] Because OpenAI has waived the additional objections in Exhibit 1, the Court should compel OpenAI to further supplement its responses without the caveats that are the result of OpenAI's untimely supplemental objections.

Much like in Kadrey v. Meta, this Court should reject OpenAI's arguments. OpenAI argues that [REDACTED]

[REDACTED] In *Kadrey*, the plaintiffs asked Meta to admit or deny that the plaintiffs' works were in the Books3 dataset. *Kadrey et. al., v. Meta Platforms, Inc.*, No. 23-CV-03417, 2024 WL 5008065 at *1 (N.D. Cal. Dec. 6, 2024). Meta responded that it lacked sufficient information to admit or deny each RFA because the plaintiffs "failed to produce the deposit copy" and admits only that "text from a published and commercially-available version of [each work] is included in the dataset commonly known as Books3." *Id.* Magistrate Judge Hixson wholly rejected Meta's arguments about deposit copies and registrations as "frivolous" because (1) after the Copyright Act of 1976, works need neither registrations nor deposit copies to have copyright protection, (2) nearly all the books at issue were written since the Copyright Act of 1976 took effect and are thus automatically copyrighted, and (3) the RFAs ask about the works, not copyright registrations or copyrights. *See id.* at *2. The Court compelled Meta to supplement its RFA responses.

The same result should obtain here. [REDACTED]

[REDACTED] the Court should compel OpenAI to admit or deny Plaintiffs' RFAs [REDACTED]

March 31, 2025

Page 3

OpenAI's [REDACTED] objection is no basis to deprive Plaintiffs of highly relevant discovery on the contents of OpenAI's training datasets. OpenAI also objects to Plaintiffs' RFAs because [REDACTED]

Ex. 1 at 5. As

discussed above, [REDACTED] are inapposite. Any burden associated with OpenAI [REDACTED]

[REDACTED] is a problem of OpenAI's own making. Plaintiffs' RFAs don't require OpenAI to [REDACTED] to respond to the RFAs, *see Kadrey*, 2024 WL 5008065 at *1, and OpenAI cannot refuse to respond to these RFAs based on its [REDACTED] objection.

OpenAI's other objections are similarly meritless. OpenAI's supplemental responses assert several additional objections, [REDACTED]

All fail. *First*, OpenAI lodges a frivolous objection [REDACTED]

[REDACTED], even though OpenAI uses these terms routinely without issue, [REDACTED]

Courts consistently overrule objections to the use of "nontechnical English words or phrases that any litigant of ordinary intelligence . . . ought to have no difficulty understanding." *U.S. Bank Nat'l Ass'n v. Triaxx Asset Mgmt. LLC*, 2020 WL 9549505, at *5 (S.D.N.Y. Nov. 30, 2020). And the Court has already decided which "Large Language Models" are in the current scope of discovery. Dkt. 293 at 2 ("Discovery and production regarding OpenAI's large language models is limited to the models identified in response to Interrogatory #11").

Second, OpenAI's [REDACTED] are unpersuasive. Plaintiffs are only asking OpenAI to respond to 290 RFAs at this stage, and courts have required parties to respond to far more RFAs. *See, e.g., Sec. & Exch. Comm'n v. Rayat*, 2022 WL 1606953, at *2 (S.D.N.Y. May 19, 2022) (1,010 RFAs). Moreover, OpenAI's refusal to respond to Plaintiffs' RFAs presents the greatest burden of all. OpenAI might take the position that Plaintiffs' books are included in the training data, or not included the training data, but foot dragging and fence sitting increases costs on both parties. Plaintiffs are trying to efficiently narrow the issues before trial, while OpenAI is trying to avoid its discovery responsibilities in order to keep its options open. *See Woodward v. Holtzman*, 329 F.R.D. 16, 26 (W.D.N.Y. 2018) (party may not avoid admissions because admissions "may prove decisive to the case").

Third, OpenAI's objection to the [REDACTED] is a red herring. The OCR copies that Plaintiffs provided were clear enough to allow OpenAI to say that [REDACTED] *See* Ex. 1 at 5. There is no reason why OpenAI cannot use these same PDFs to admit or deny the questions that Plaintiffs asked.

For the reasons described above, the Court should compel OpenAI to further supplement its RFA responses.

¹ OpenAI's Fine-tuning webpage uses the word "train" or other variations (training, pre-training, etc.) 81 times. *See* <https://platform.openai.com/docs/guides/fine-tuning> (visited March 30, 2025).

March 31, 2025

Page 4

Sincerely,

LIEFF CABRASER HEIMANN SUSMAN GODFREY LLP COWAN, DEBAETS,
& BERNSTEIN LLP ABRAHAMS & SHEPPARD
LLP

/s/ Rachel Geman

Rachel Geman

/s/ Rohit Nath

Rohit Nath

/s/ Scott J. Sholder

Scott J. Sholder